

# N471 Project Milestone 1 Proposal CARLSON

Nicole Carlson

February 4, 2018

## Overview & Motivation

email: "[nicole.carlson@emory.edu](mailto:nicole.carlson@emory.edu)"

N471 Final Project Title: "BMI-, Parity-, and Mode of Labor Onset- Individualized Partographs from Consortium on Safe Labor dataset"

Repository for this project planning can be found at [this github location](#).

This project is part of my NIH-funded K01 study, "Metabolomics of Labor Dysfunction in African-American Women". Partographs are graphical representations of cervical dilation over time during labor. Traditionally, partographs include representations (lines, colored areas, etc) of population-generalizable limits for normal labor progress. For my new partographs, I will be using data from a large, multi-site study (the Consortium on Safe Labor, or CSL study, N=228,000) to show normal labor progress for different groups of women. Multiple partographs will be created for (individualized by) BMI grouping, parity (first labor vs. not) and by mode of labor onset (spontaneous labor vs. induced labor). Eventually, I will be using these individualized partographs to identify women in my study who demonstrated labor dystocia, defined as labor progression that is abnormally slow.

For this N741 project, I propose to use Rmd to clean the CSL dataset, select samples based on maternal characteristics (BMI group, parity, mode of labor onset), and prepare for multivariate analyses that will ultimately produce the labor traverse times I will need to build the individualized partographs. I've started this process in SPSS, but would like to instead do the work in Rmd format.

I am motivated to do this work for my N741 project out of my desire to force myself to competency in using R and Rmd. I completed another large clinical database secondary analysis during my dissertation, and saw first-hand how easy it was to make mistakes that cost considerable time and effort when I was not easily able to reproduce my research. In addition, I have several future analyses planned with the CSL dataset, and the work I expend here in cleaning and setting up groups with the CSL will be useful later.

## Project Objectives

1. Download dataset into R and set up variable properties and lables. Use Rmd and git (private repository) to store data, version control for manipulations.
2. Clean dataset in R, using Rmd to demonstrate and make possible reproducible research cleaning.
3. Create groups from cleaned dataset--a total of 24 groups:

- Parity: Nulliparous or Multiparous
- BMI group at the time of labor admission
  - BMI  $\leq 18.99$
  - BMI 19.00-24.99
  - BMI 25.00-29.99
  - BMI 30.00-34.99
  - BMI 35.00-39.99
  - BMI  $\geq 40.00$
- Mode of birth: Spontaneous or Induced
- 4. Link repeated measures data (sequential cervical exams and oxytocin dose data over time) to subject data. *(optional, additional project depending on time for Objectives 1-3)*
- 5. Perform multivariate analyses to calculate traverse times for each grouping. *(optional for N741 project, depending on time for Objectives 1-4)*

## Data

I received a copy of the CSL dataset under a DUA from NICHD for the purposes of completing this project. More information about the dataset can be found [here](#).

## Data Wrangling

*This project will involve a moderate amount of data wrangling:*

**Data cleaning**-will need to clean dataset for cases of missing BMI. BMI is the major variable I will be using to create groups, yet it is missing in about 30% of cases from CSL. I will also check for missing cases on other demographic variables (race, health insurance), plus other variables I will use to create groups (height, weight at labor admission, parity, mode of labor onset), and will delete cases missing these variables. In addition, I will check for missing cases on variables I will use as exclusions (women who did not labor, non-vertex presentations, non-term gestations, non-singleton gestations), and delete those that are missing. Although there will be many irrelevant variables (for this project) in CSL, I will likely keep most of these variables in my cleaned dataset, for future analysis.

**Data extraction**-exclusions will be applied, including those for non-labor, non-vertex presentations, non-term gestations, non-singleton gestations.

**Data reshaping**-calculate BMI, parity, labor onset variables from CSL variables. For objective 4, change dataset to long-form, merge with repeated measures dataset.

## Exploratory Analysis

I will explore my dataset by producing a table showing demographic variables (age, race, ethnicity, marital status, type of health insurance) and variables I will use to create each group (delivery BMI, parity, mode of labor onset), as well as a few of the most fundamental labor outcomes (mode of birth, labor anesthesia/analgesia, maternal and neonatal

mortality, maternal and neonatal morbidity) in the cleaned dataset with all exclusions applied.

I've had a start on using R to build this type of summary table using the `table_summary` function. I will expand this build to include all of the variables listed above in the table. Then, after I've created groups by BMI, parity, and mode of birth, I can repeat the summary stats function on each group to see how these basic demographic and labor outcomes might be different by group.

Here is the format for the basic summary table I will produce for each group:

Variable	#(%) or median (IQR)
Maternal age	
Race	
Marital status	
Health Ins	
Gestational age	
Mode labor onset	
Labor Analgesia	
Mode of birth	
Maternal death	
Hemorrhage	
3/4th laceration	
Shoulder dystocia	
Neonatal death	
Apgar<7 @ 5 min	
NICU admission	
Neonatal seizure	

For the entire sample, I would additionally include in this table the variables used to build groups, including BMI, parity, and mode of labor onset. Finally, as part of my exploratory analysis, I may create some visualizations on particular variables. I will first examine the summary statistic tables to decide which variables I might want to visualize.

## Analysis

This last section of the project will depend on the amount of time I have remaining after completing Objectives 1-3. Assuming I have time after applying exclusion criteria, creating groups, and conducting exploratory analysis, the next step in this project will be merging the long-form CSL dataset, which contain basic information on each labor, to the wide-form repeated-measures dataset, which contains the time and results for each cervical examination. I plan to do this conversion using the `gather()` function, which is part of the `tidyr` package. I've found some basic instructions for this operation on [this github tutorial](#).

If I am successful in merging the repeated measures information and converting to a wide-form dataset, I can then conduct a survival analysis that will produce the traverse times for women in different groups to change cervically from 4-5cm, 5-6cm, 6-7cm, 7-8cm, 8-9cm, and 9-10cm. As part of my K01, Dr. Paul will help answer questions I might have about conducting survival analysis, once I get to this section.

### **Proposed Timeline for Carlson project N741**

Date	Completed component of project
2/21/18	Objective 1 complete. Begin data wrangling.
2/28/18	Objective 2 complete.
3/07/18	Objective 3 complete.
3/15/18	Create tables/visulatizations of Exploratory Analyses. Milestone 2 submitted, submit working prototype
3/29/18	Merge long-form with repeated measures CSL datasets
4/11/18	Survival analyses complete. Write up project notes.
4/19/18	Milestone 3, submit final PDF of project.
4/26/18	Submit final presentation material